Correlation Practice Problems

1. Consider the following data set:

x	5	6	7	7	8	8	8	8	9	9	10	10	11	11	12	12
y	4.2	5	5.2	5.9	6	6.2	6.1	6.9	7.2	8	8.3	7.4	8.4	7.8	8.5	9.5

(a) Draw a scatter diagram with the *x*-axis starting at 0 and ending at 30 and with the *y*-axis starting at 0 and ending at 20.
(b) Compute the linear correlation coefficient.
(c) Now multiply both *x* and *y* by 2.
(d) Draw a scatter diagram of the new data with the *x*-axis starting at 0 and ending at 30 and with the *y*-axis starting at 0 and ending at 20. Compare the scatter diagrams.
(e) Compute the linear correlation coefficient.
(f) Conclude that multiplying each value in the data set by a nonzero constant does not affect the correlation between the variables. Explain why this is the case.

2. Lyme disease is an inflammatory disease that results in a skin rash and flulike symptoms. It is transmitted through the bite of an infected deer tick. The following data represent the number of reported cases of Lyme disease and the number of drowning deaths for a rural county in the United States.

Month	**J**	**F**	**M**	**A**	**M**	**J**	**J**	**A**	**S**	**O**	**N**	**D**
Cases of Lyme Disease	3	2	2	4	5	15	22	13	6	5	4	1
Drowning Deaths	0	1	2	1	2	9	16	5	3	3	1	0

(a) Draw a scatter diagram of the data using cases of Lyme disease as the explanatory variable.
(b) Compute the correlation coefficient for the data.
(c) Based on your results from parts (a) and (b), what type of relation exists between the number of reported cases of Lyme disease and drowning deaths? Do you believe that an increase in cases of Lyme disease causes an increase in drowning deaths? What is a likely lurking variable between cases of Lyme disease and drowning deaths?

3. Researchers wondered whether the size of a person's brain was related to the individual's mental capacity. They selected a sample of right-handed introductory psychology students who had SAT scores higher than 1350. The subjects took the Wechsler Adult Intelligence Scale-Revised to obtain their IQ scores. MRI scans were performed at the same facility for the subjects. The scans consisted of 18 horizontal MR images. The computer counted all pixels with a nonzero gray scale in each of the 18 images, and the total count served as an index for brain size.

Gender	**MRI Count**	**IQ**
Female	816,932	133
Female	951,545	137
Female	991,305	138
Female	833,868	132
Female	856,472	140
Female	852,244	132
Female	790,619	135
Female	866,662	130
Female	857,782	133
Female	948,066	133
Male	949,395	140
Male	1,001,121	140
Male	1,038,437	139
Male	965,353	133

Gender	MRI Count	IQ
Male	955,466	133
Male	1,079,549	141
Male	924,059	135
Male	955,003	139
Male	935,494	141
Male	949,589	144

Source: L. Willerman, R. Schultz, J. N. Rutledge, and E. Bigler (1991). "In Vivo Brain Size and Intelligence," *Intelligence*, 15, 223–228.

(a) Draw a scatter diagram treating MRI count as the explanatory variable and IQ as the response variable. Comment on what you see.
(b) Compute the linear correlation coefficient between MRI count and IQ. Are MRI count and IQ linearly related?
(c) A lurking variable in the analysis is gender. Draw a scatter diagram treating MRI count as the explanatory variable and IQ as the response variable, but use a different plotting symbol for each gender. For example, use a circle for males and a triangle for females. What do you notice?
(d) Compute the linear correlation coefficient between MRI count and IQ for females. Compute the linear correlation coefficient between MRI count and IQ for males. Are MRI count and IQ linearly related? What is the moral?

4. The following data represent the number of licensed drivers in various age groups and the number of fatal accidents within the age group by gender.

Age	Number of Male Licensed Drivers (000s)	Number of Fatal Crashes	Number of Female Licensed Drivers (000s)	Number of Fatal Crashes
<16<16 less than 16	12	227	12	77
16–20	6,424	5,180	6,139	2,113
21–24	6,941	5,016	6,816	1,531
25–34	18,068	8,595	17,664	2,780
35–44	20,406	7,990	20,063	2,742
45–54	19,898	7,118	19,984	2,285
55–64	14,340	4,527	14,441	1,514
65–74	8,194	2,274	8,400	938
>74>74 greater than 74	4,803	2,022	5,375	980

Source: National Highway and Traffic Safety Institute

(a) On the same graph, draw a scatter diagram for both males and females. Be sure to use a different plotting symbol for each group. For example, use a square $(\blacksquare)$ for males and a plus sign $(+)$ for females. Treat number of licensed drivers as the explanatory variable.
(b) Based on the scatter diagrams, do you think that insurance companies are justified in charging different insurance rates for males and females? Why?
(c) Compute the linear correlation coefficient between number of licensed drivers and number of fatal crashes for males.
(d) Compute the linear correlation coefficient between number of licensed drivers and number of fatal crashes for females.
(e) Which gender has the stronger linear relation between number of licensed drivers and number of fatal crashes. Why?